DOCUMENT RESUME

ED 373 088 TM 021 977

AUTHOR Evans, John Andrew; Ackerman, Terry
TITLE Diagnostic Value Resulting from the IRT Modeling of IGAP Reading Data: Using A Graded Response Model To Retrieve and Utilize More Information.
PUB DATE Apr 94
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994). For related documents, see TM 021 975-976.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Ability; *Educational Diagnosis; Elementary Education; Elementary School Students; Information Retrieval; *Item Response Theory; Models; Reading Achievement; *Reading Tests; Research Methodology; Test Format; *Test Items
IDENTIFIERS *Graded Response Model; Illinois; *Illinois Goal Assessment Program

ABSTRACT

The strengths of item-response theory (IRT) are used to examine the degree of information individual test items provide, as well as to investigate how the individual item types contribute to the overall measurement accuracy of the Illinois Goal Assessment Program (IGAP) reading test. Using the graded-response model of Samejima (1969), the amount of information each subtest (narrative and expository) provides about the underlying latent ability is studied. Where an item type provides the most information along this ability scale, and how the different item formats (e.g., number of correct inferences) differ in terms of ability to discriminate between levels of reading proficiency are also studied. Data sets of 4,837, 4,840, and 5,011 randomly selected examinees were obtained for grades 3, 6, and 8, respectively. While the expository subtest is generally more informative than the narrative subtest across the three grade levels for low to moderate theta values, the difference does not appear to be substantial. The graded response model appears to be a promising tool that allows examination of the information from each subtest. Fourteen figures illustrate the findings. (Contains 10 references. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 373 088

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TERRY ACKERMAN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Diagnostic Value Resulting from the IRT Modeling of IGAP Reading Data: Using a Graded Response Model to Retrieve and Utilize More Information

John Andrew Evans
Terry Ackerman

Running head: IRT Modeling of IGAP Reading Data

Paper presented at the 1994 AERA Annual Meeting, New Orleans

TM021977

BEST COPY AVAILABLE

Abstract

The purpose of this inquiry is to utilize the strengths of Item Response Theory (IRT), to examine the degree of information individual test items provide, as well as to investigate how the individual item types contribute to the overall measurement accuracy of the IGAP reading test. Using Samejima's (1969) graded response model, this paper compares the amount of information each subtest, narrative and expository, provides about the underlying latent ability; where along this ability scale an item type provides the most information; and how the different item formats (e.g., number of correct inferences) differ in terms of their ability to discriminate between levels of reading proficiency.

Introduction

In contrast to more traditional reading assessments that use isolated paragraphs and fragmented text, the passages used in the Illinois Goal Assessment Program (IGAP) are intact pieces of literature, stories, and essays that match classroom reading assignments and typical student reading experiences. The items associated with each passage require students to demonstrate various levels of cognitive skills, from explicit response to drawing conclusions that are not directly stated, solving problems not discussed within the text, and using information derived from the reading passage. Because texts often support more than one correct inference, the Illinois reading assessment uses a multiple response (or multiple correct) rather than a multiple choice format. There may be more than one correct conclusion or inference to each item with credit awarded each time a correct or incorrect inference is selected.

"IGAP defines reading as a dynamic process by which readers combine background knowledge, reading ability, strategic awareness, and information from the text to construct meaning."[1] The format to assess an examinee's ability to construct meaning presents the examinee with two passages, one narrative (story type) and one expository (informational type) with questions accompanying each passage. The test is administered in two 40-minute sessions, one type of passage per session, with a rest period between sessions.

The item structure utilizes a multiple correct inference format where each item may have one, two, or three correct inferences. Thus, examinees have to identify which conclusions are both correct and

incorrect. Each sub-test (narrative and expository) has 15 items. Each item is constructed using one or more of the five questioning types: Explicit, Inference Level I, Inference Level II, Application Transfer, or Vocabulary. Third grade examinees respond directly on the test booklet, while sixth and eighth grade examinees mark a separate answer sheet.

Results from the IGAP reading examination are used to identify whether an examinee fails to meet, meets, or exceeds a predetermined standards. Two cut scores along the underlying ability scale differentiate between fails to meet and meets the standard, and meets and exceeds the standard.

This examination structure would appear to be rich in information necessary to make critical attainment decisions. One of the strengths of Item Response Theory (IRT) modeling, when all of the underlying assumptions of the model are met, is that practitioners can gain a great deal of information about individual test items and how they contribute to the overall measurement accuracy of the test.

A review of reading assessment literature produced no study that applied a graded response model to analyze reading test results. However, within the area of language arts, a partial credit model has been utilized to analyze narrative writing tasks in order to identify aspects of writing that function differently (Harris, Laan, and Mossen, 1988). Additionally, from an IRT perspective, Ackerman (1986) used a graded model to compare holistically scored essays with multiple choice writing tests in an attempt to see which is more informative, and at which abilities the most information is provided.

Although the partial credit model requires a continuum, the graded model requires only an ordering. Items within the IGAP reading test are not structured to be interpreted on a continuum, but do allow ordering.

Graded Response

Prior to the 1993 IGAP reading examination, test results had been analyzed and equated by procedures that are rooted in classical test theory and limited to free response items that had been dichotomously scored. Unlike such binary items, the item variable scale in the graded scoring procedure is divided into ordered categories. As such, the lowest category contributes least to a person's test score while the highest category would contribute most.

It is an underlying assumption that an examinee's response to an item scored on a graded basis possesses a hypothetical continuous item variable ranging from $-\infty$ to $+\infty$ and has been divided into m response categories for a given item. The response categories are ordered, with k denoting an arbitrary category, k=0,1, 2,...m_i, where m_i, is the number of response categories for item i (Baker, 1992).

The probability of an examinee responding to category k or higher can be denoted as:

$$P_{(x=k)} = \frac{1}{1 + \exp[-a(\theta - b_{k-1})]} - \frac{1}{1 + \exp[-a(\theta - b_k)]}$$

where b_k is the difficulty level from category k (Hambleton, Swaminathan, & Rogers, 1991).

The IGAP reading examination provides six categories, k=0,1,...5. The assumption of equal discrimination parameters for k=0,1,...5, that is, the homogeneous case of the graded response model, yields k=0 and k=5 as monotonic and k=1,2,3,4 as generally non-monotonic (Hambleton & Swaminathon, 1985). When plotted in concert (Figure 1), the interrelationship among the six response categories is more easily recognized.

Insert Figure 1 about here

At the higher θ levels, category k=5 has the highest probability of occurring; at the lowest θ levels category k=0 has the highest probability of occurring. For the middle ability levels, categories k=1,2,3,4, the probability values are most probable.

At each grade level measured, the IGAP reading examination has 15 items with a 6-point scale (0, 1, 2, 3, 4, and 5) for both the narrative and expository passages,. For this analysis (15 x 5) + 15 = 90 item parameter values would be estimated for both sub-tests (narrative and expository) , using the graded response model in MULTILOG (Thissen, 1991). The program employs MMLE to obtain item parameter estimates.

Item parameters estimated to fit the graded response model were subsequently used to compute the item information functions. The amount of information yielded by an item at ability level θ in the polytomous case, can be expressed as:

$$I_i(\theta) = \sum_{k=1}^{m} I_k(\theta) P_k(\theta)$$

where the quantity $I_k(\theta)\ P_k(\theta)$ is the amount of information share of category *k* (Baker, 1991, p.240).

The amount of item information yielded by polytomous scoring when compared to dichotomous scoring will result in an increase in the amount of item information (Samejima, 1969, p.40). As such, the graded response case would be expected to produce a smaller standard error for the estimate of the examinee's latent ability than the dichotomous case (Baker, 1991, p. 244).

Comparing information functions for the two components (narrative and expository) of the IGAP reading examination that are measuring the same ability, θ (Bolt & Ackerman, 1994) can be written as:

$$RE(\theta) = \frac{I_N(\theta)}{I_E(\theta)}$$

where $RE(\theta)$ is the relative efficiency and $I_N(\theta)$ and $I_E(\theta)$ are the information functions for the narrative subtest and the expository subtest, defined over a common ability scale (θ) (Hambleton, et al., 1991, p.96).

Method

Data sets of 4537, 4840, and 5011 randomly selected examinees were obtained for grades three, six, and eight, respectively. The data set consisted of response patterns for the fifteen testlets within both the narrative and expository subtests; a total of thirty items. Each of the thirty items was scored polytomously with *k* categories, *k*=0,2,...5. Item

parameters were estimated using MULTILOG (Thissen, 1991). FORTRAN programs were constructed to produce item and test information functions, as well as relative efficiency comparisons between the narrative and expository subtests at each grade level.

Results

Comparing sub-tests

Figures 2-4 plots contain the relative efficiency of the narrative and expository subtests of grades three, six and eighth respectively. Figure 2 illustrates that the expository subtest provides more information for ability levels less than a θ value of +2.00. The narrative subtest provides greater information for those examinees with a θ value greater than +2.00. But, in both cases the extent of "more" information is less than ten percent. At grade six, the amount of information provided by a specific sub-test varies across the ability scale. For those examinees with extremely low θ values (less than -2.75) and moderately high θ values (greater than +1.6) The narrative subtest provides more information. For those sixth grade examinees with θ values between -1.7 and +1.6, the expository subtest is more informative. Again, as was the case in grade three, the extent of greater information is less than ten percent.

Insert Figures 2-4 about here

The amount of information provided by each sub-test across the common ability scale is easily observed in Figure 3. Figure 4 illustrates that the Expository Sub-test is increasingly more informative as one moves up

the ability scale, providing greater than ten percent more information for those examinees with estimated θ values greater than +1.5.

As a practical matter, the test constructor(s) can observe the relative efficiency of one sub-test compared with another throughout the θ range. For example, at θ value 0.5 the eighth grade expository subtest is providing approximately 7% more information than the eighth grade narrative subtest. That is, to have an equal amount of information provided by each sub-test, the test constructor(s) would need to increase the number of Narrative items by 1.

Comparing questioning types

Each five part testlet can be denoted by the five questioning formats, Explicit, Inferential I, Inferential II, Application Transfer, and Vocabulary. The testlet may consist of a single item type across all five questions or may consist of multiple item types across all five parts. For this analysis, item types across testlets was selected. That is, each item type (Explicit, Inferential I, Inferential II, Application Transfer, and Vocabulary) is represented within the analysis.

Figures 5-7 represents the item information function by number of correct inferences. Regardless of question type, more information is provided for examinees of estimated lower ability. In general, the amount of information declines for θ values of 1.00 regardless of grade. It should be noted that the grade eight exam possesses only four questioning types, as one of the questioning types is not included in the examination.

Insert Figures 5-7 about here

Comparing item by number of correct inferences

To compare items of different number of correct inferences (i.e., one, two, or three), their information functions are examined. The findings are similar to those seen when information functions for item types are compared. Regardless of grade level, items with one, two, or three correct responses provide the most information for examinees of with estimated θ values less than 0 (Figures 8-10).

Insert Figures 8-10 about here

Figures 11 graphically represent the mean item information function by number of correct inferences. The mean item information function accounts for the differing number of items with one, two, or three correct. responses. The testlet information function is the sum of the item information functions. Without compensating for the number of items within a testlet, one may assume a given item type is more nformative because its sum is larger when , in fact, its sum is larger because there are more items in the testlet.

For grades three, six, and eight the one correct response items provide more information than the two correct response items, and both provide more information than the three correct response items.

Insert Figures 11 about here

As expected, the mean item information function by number correct item type is representative of the individual items, providing more information for low to moderate θ values.

Discussion

While the expository subtest is generally more informative than the narrative subtest across the three grade levels for low to moderate θ values the difference does not appear to be substantial. If one were to try to equalize the amount of information for both sub-tests, one would need to weigh the cost of increasing the number of expository items (or decreasing the number of narrative items) and the subsequent effect of altered test length on the administrative time of the exam.

Given that one of the purposes of this exam is to identify examinees that fail to meet, meet, and exceed a predetermined standard, one would expect the items to be providing a significant amount of information at the two cut scores. IGAP reading exam items appear to provide the most information at the low ability levels, and not at either of the two cut scores (Figure 12). Few items are very informative at the moderate and high ability levels.

Insert Figure 12 about here

IRT analysis provides the IGAP reading test constructors with the opportunity to reconstruct or add items to provide information about

examinees at moderate and higher θ values while maintaining a balance between the Narrative and Expository subtests information functions.

The format of the IGAP reading examination represents an alternative to traditional reading assessment instruments, and the opportunity to provide educators with greater information regarding their students' reading performance. The graded response model (Samejima, 1969) appears to be a promising tool that allows test constructors an opportunity to investigate and compare the amount of information for each subtest (narrative and expository) and item (type and number correct) structure.

Figure Captions

Figure 1. Equal discrimination parameters.

Figure 2. Third grade narrative vs. expository relative efficiency.

Figure 3. Sixth grade narrative vs. expository relative efficiency.

Figure 4. Eighth grade narrative vs. expository relative efficiency.

Figure 5. Third grade item information function by question type.

Figure 6. Sixth grade item information function by question type.

Figure 7. Eighth grade item information function by question type.

Figure 8. Third grade item information function by number of correct responses.

Figure 9. Third grade item information function by number of correct responses.

Figure 10. Third grade item information function by number of correct responses.

Figure 11. Mean item information function by number of correct responses and grade.

Figure 12. Sixth grade mean item information function by number of correct responses.

Figure 13. Eighth grade mean item information function by number of correct responses.

Figure 14. Sixth grade item information functions with cut scores represented.

References

Ackerman, T. A.(1986). Use of the graded response IRT model to assess the reliability of direct and indirect measures of writing assessment. Mid-Western Edcuational Research Association annual meeting: Chicago, IL.

Bolt, D. & Ackerman, T. A.(1994). An examination of the influence of expository and narrative passages on the dimensionality of the I GAP reading test. AERA annual conference: New Orleans, LA.

Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). MMSS fundamentals of item response theory. Newbury Park, CA: Sage.

Hambleton, R. K. & Swaminathan, H. (1985). Item response theory principals and applications. Boston, MA: Kluwer Nijhoff

Harris. J., Laan, S. & Mossenson L. (1988). Applying partial credit analysis to the construction of narrative writing tests. Vol. 1, No. 4, Applied Measurement in Eduction.

Illinois State Board of Education (1989). The Illinois reading assessment: Classroom connections. Springfield, IL: Illinois State Board of Education.

Illinois State Board of Education (1992). The Illinois goal assessment program information bulletin. Springfield, IL: Illinois State Board of Education.

Illinois State Board of Education (1992). The Illinois goal assessment program: Reading. Springfield, IL: Illinois State Board of Education.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Eribaum.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores (No. 17). Psychometric Monograph.

Footnotes

[1]Illinois State Board of Education (1989). The Illinois reading assessment: Classroom connections. Springfield, IL: Illinois State Board of Education.

Figure 1

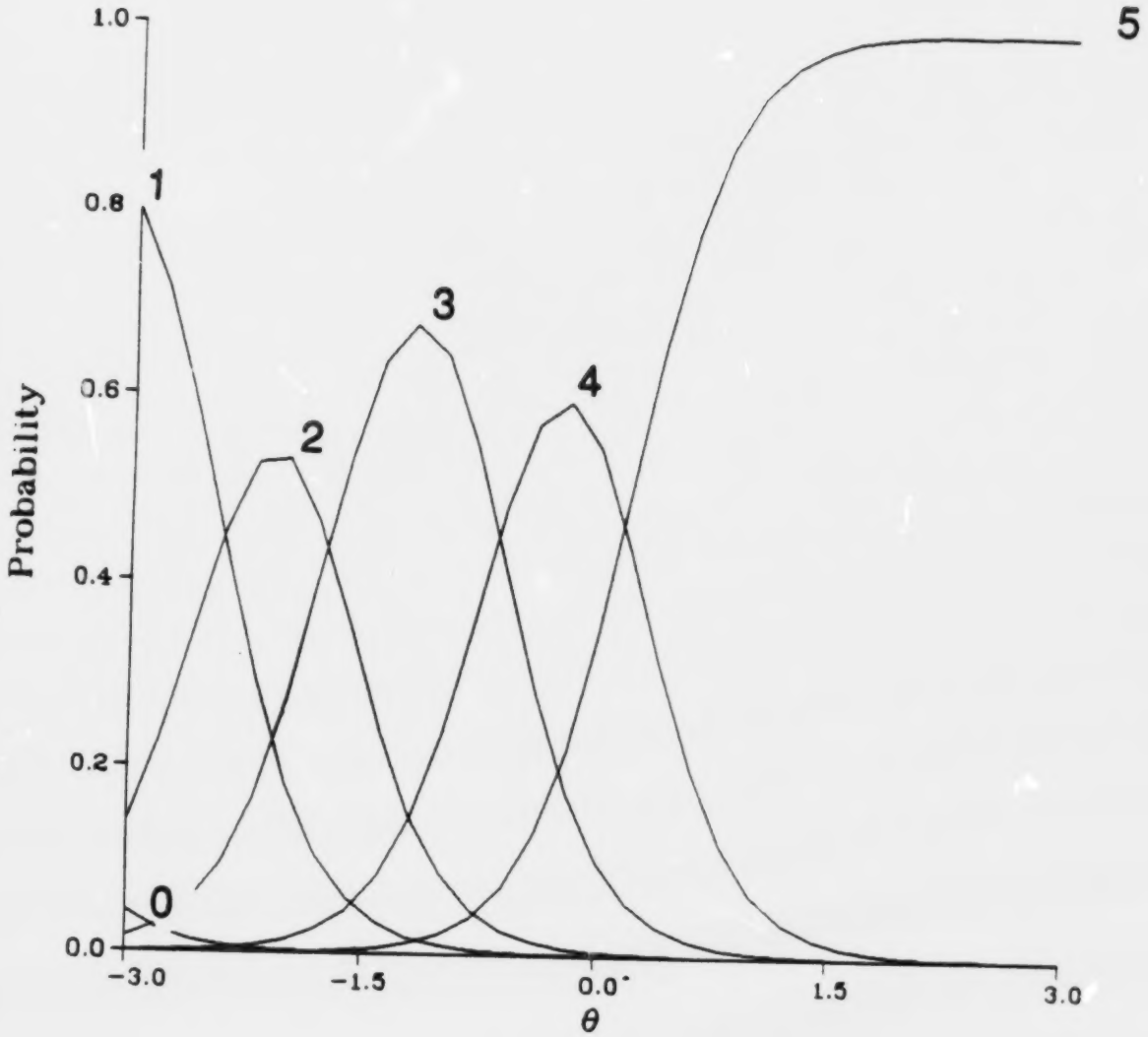


Figure 2

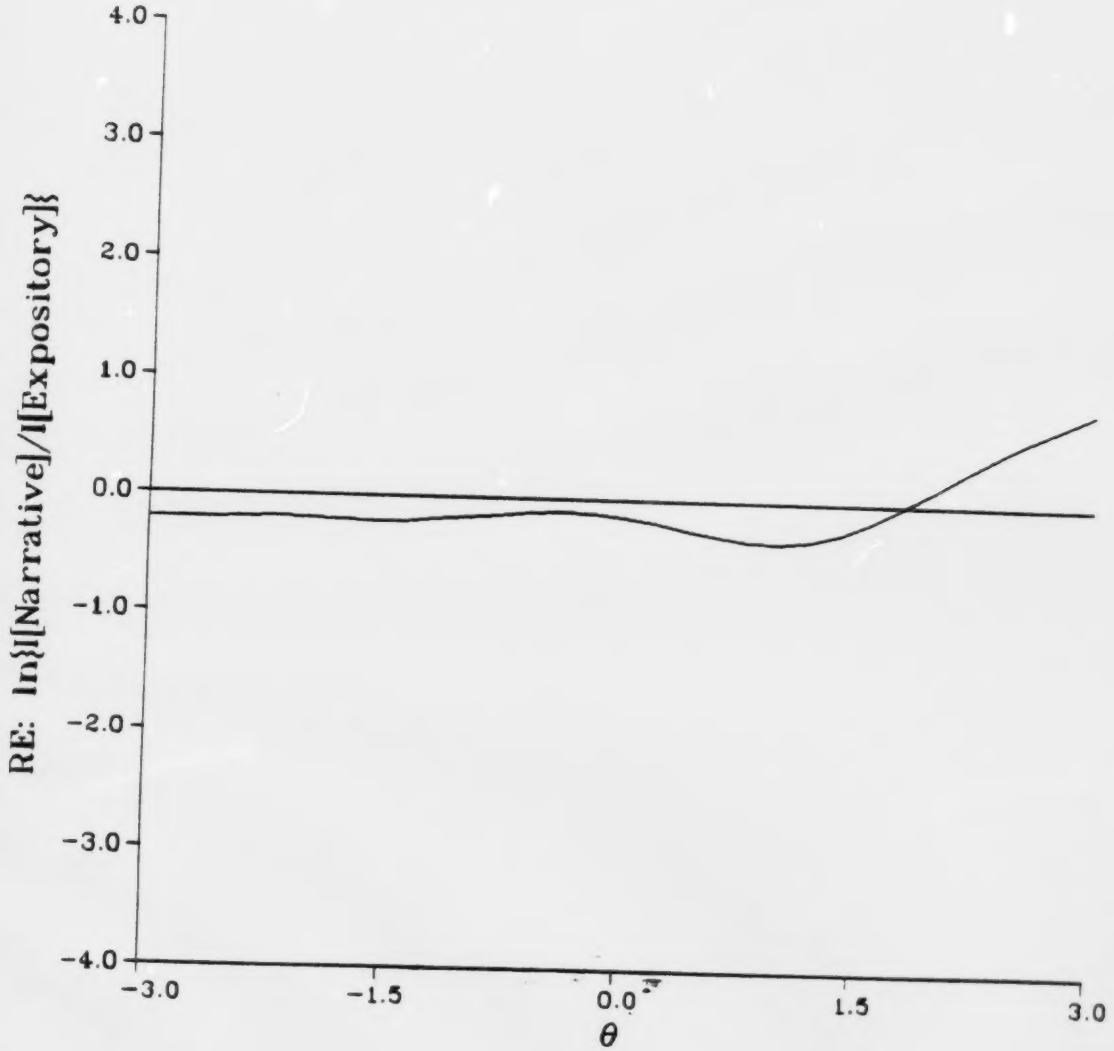

4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
RE: ln{I[Narrative]/I[Expository]}
-3.0
-1.5
0.0
1.5
3.0
θ

Figure 3

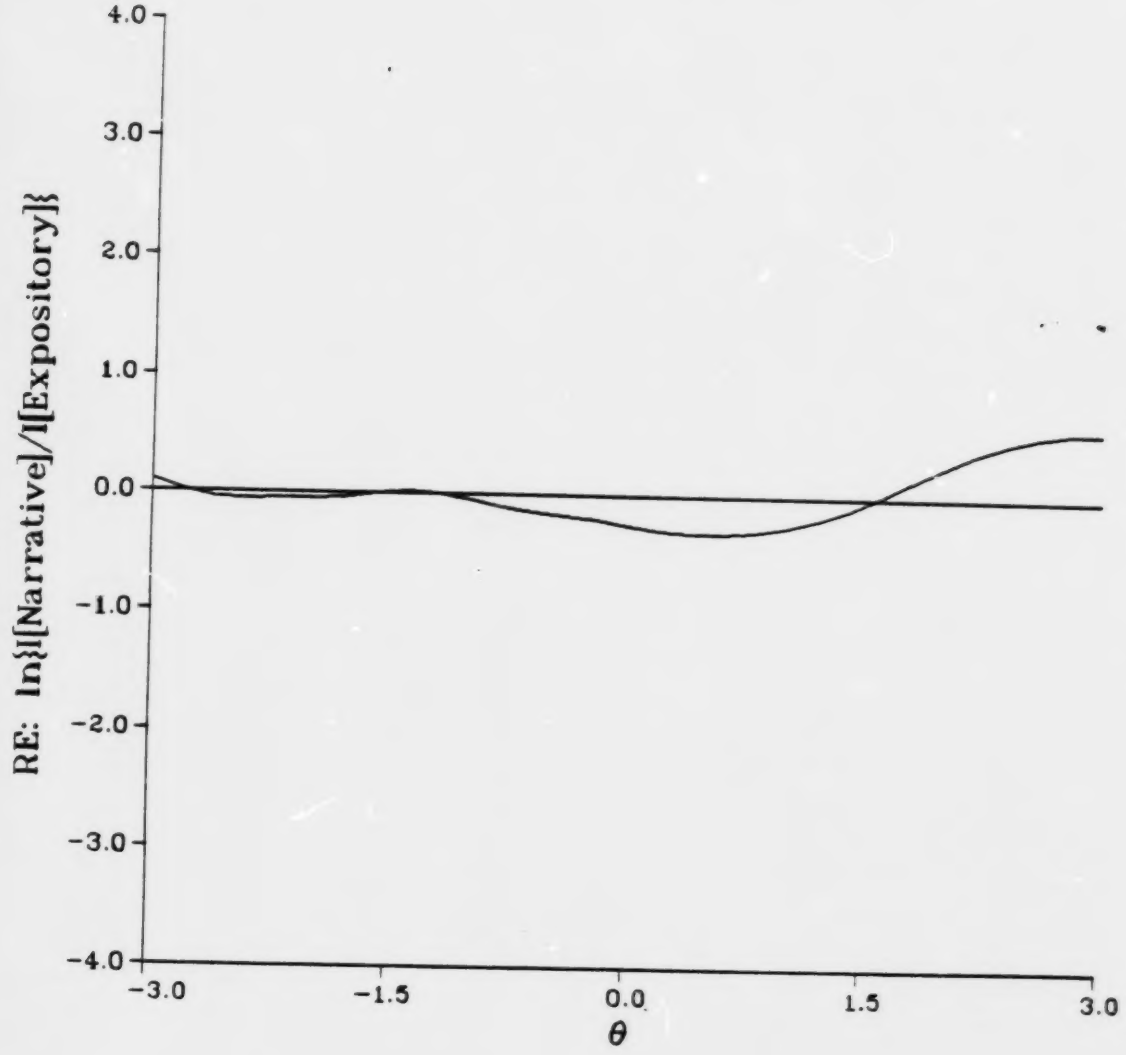

RE: ln{I[Narrative]/I[Expository]}
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-3.0
-1.5
0.0
1.5
3.0
θ

Figure 4

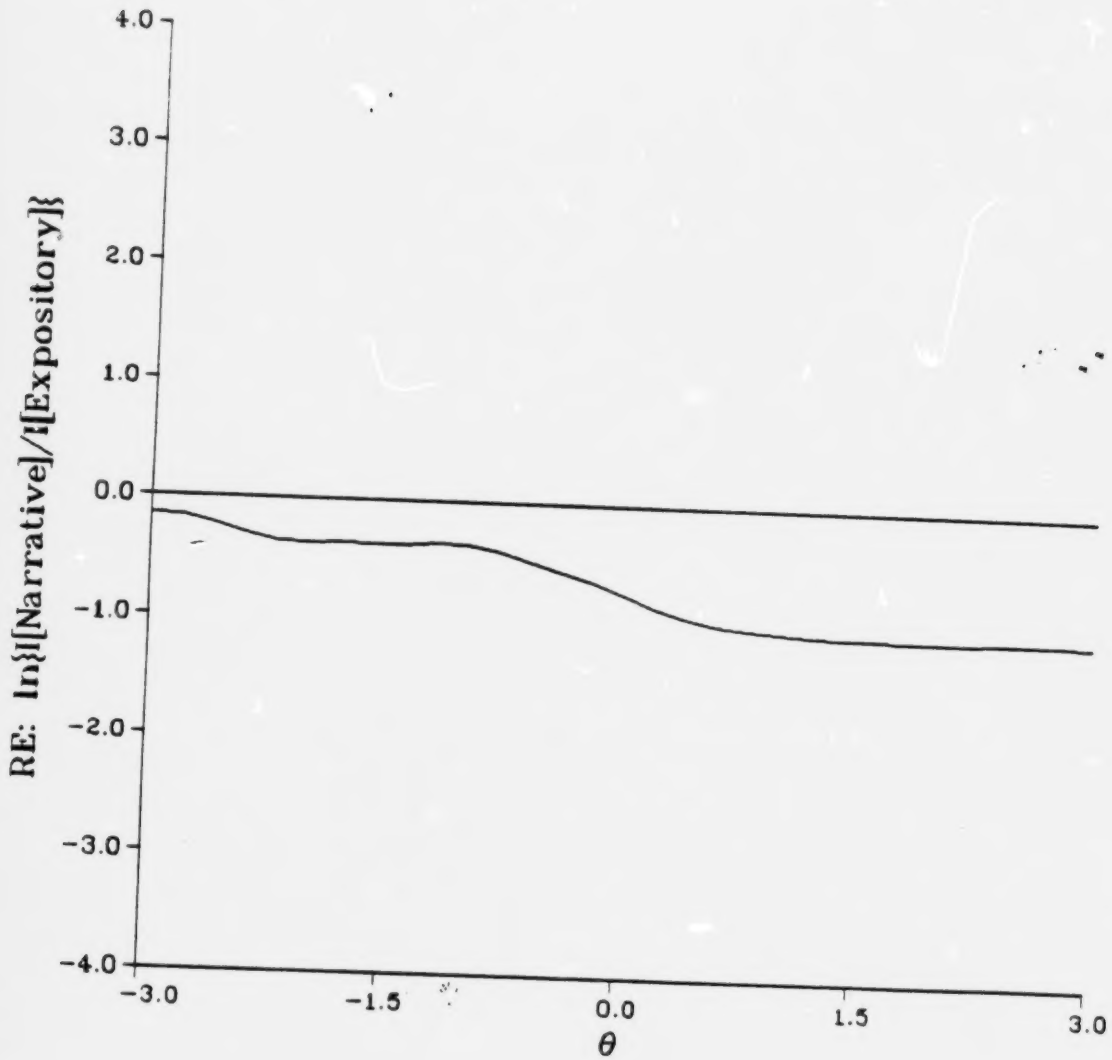

4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
RE: ln{I[Narrative]/I[Expository]}
-3.0
-1.5
0.0
1.5
3.0
θ

Figure 5

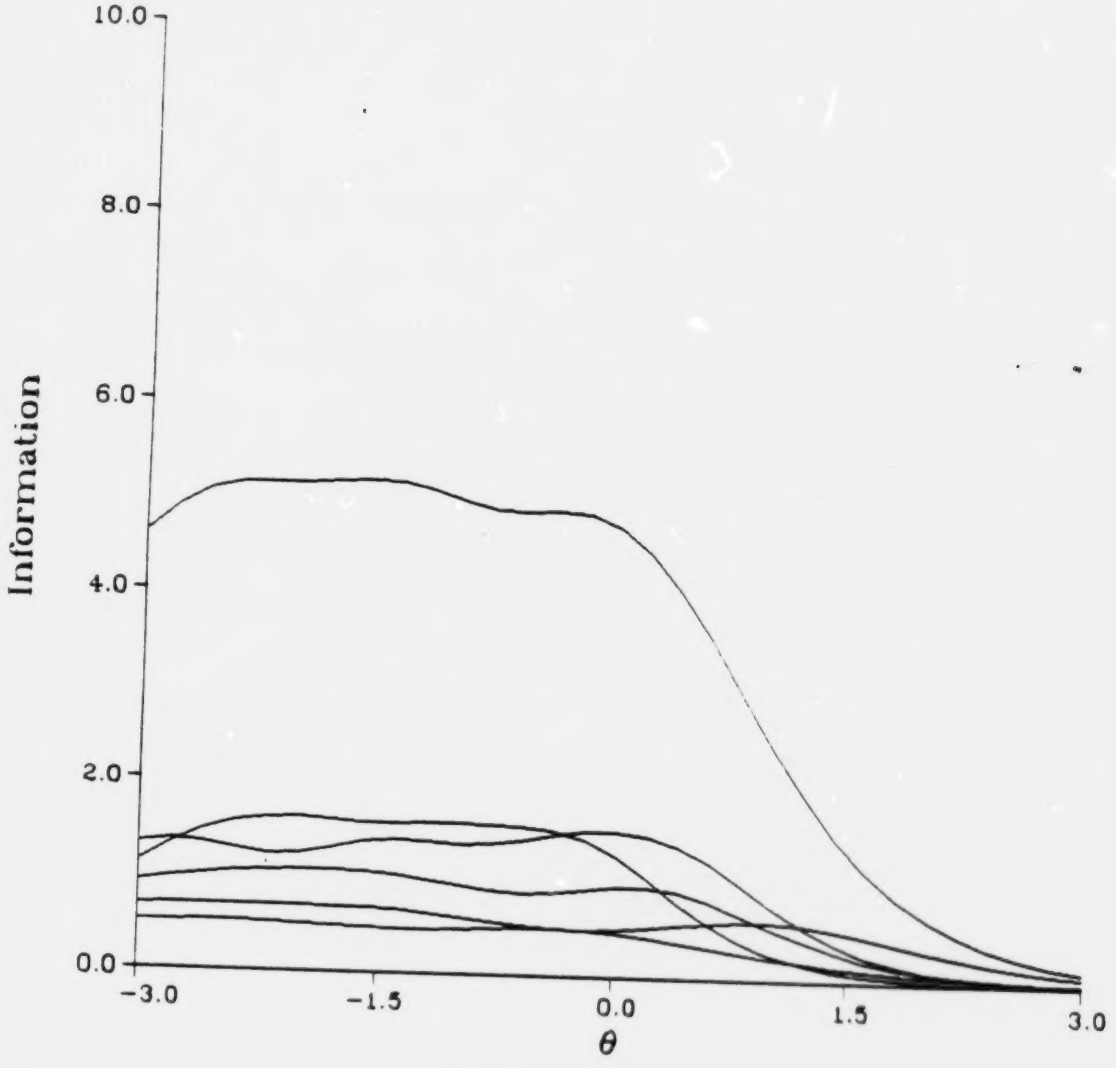


Figure 6

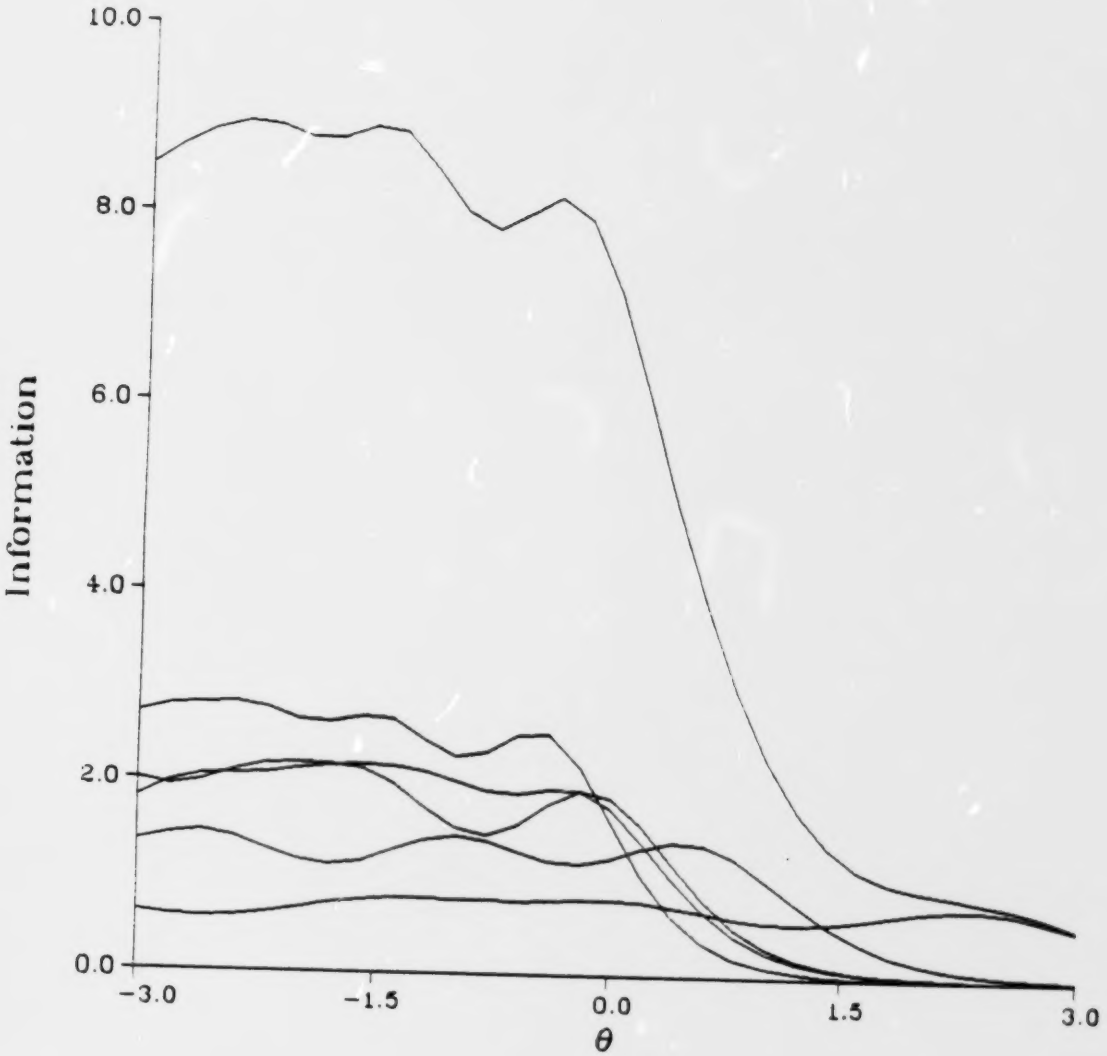


Figure 7

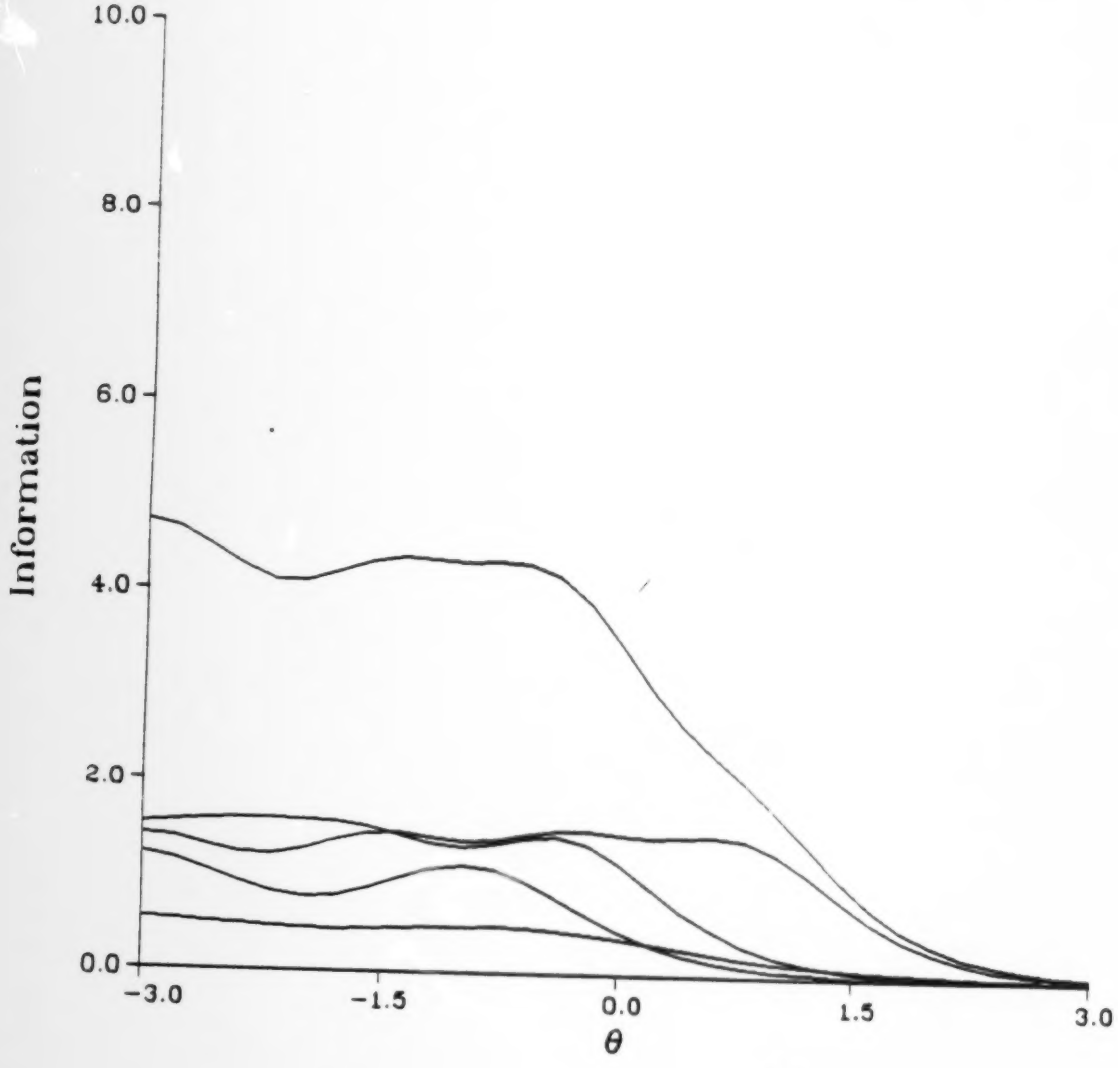
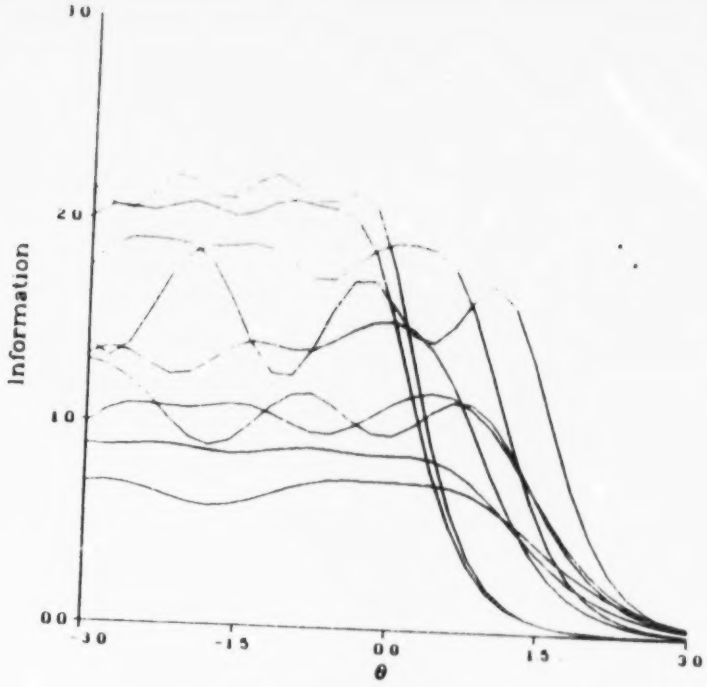
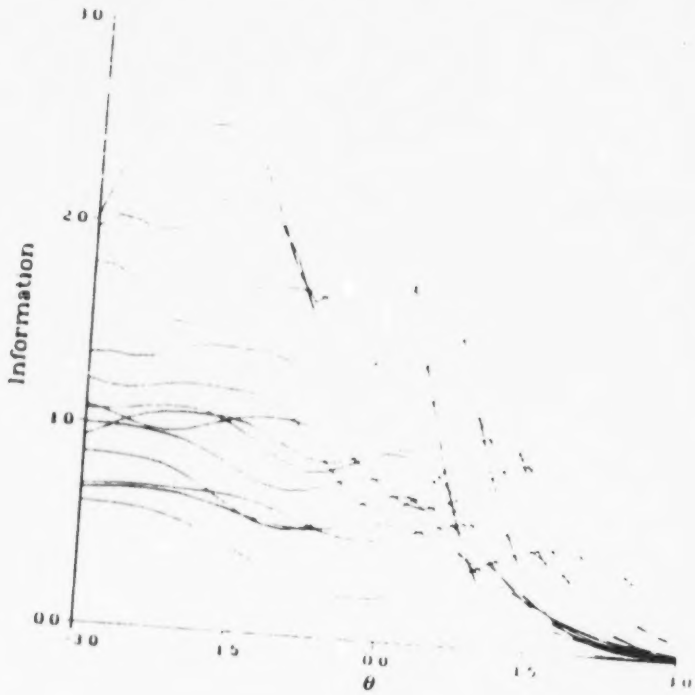


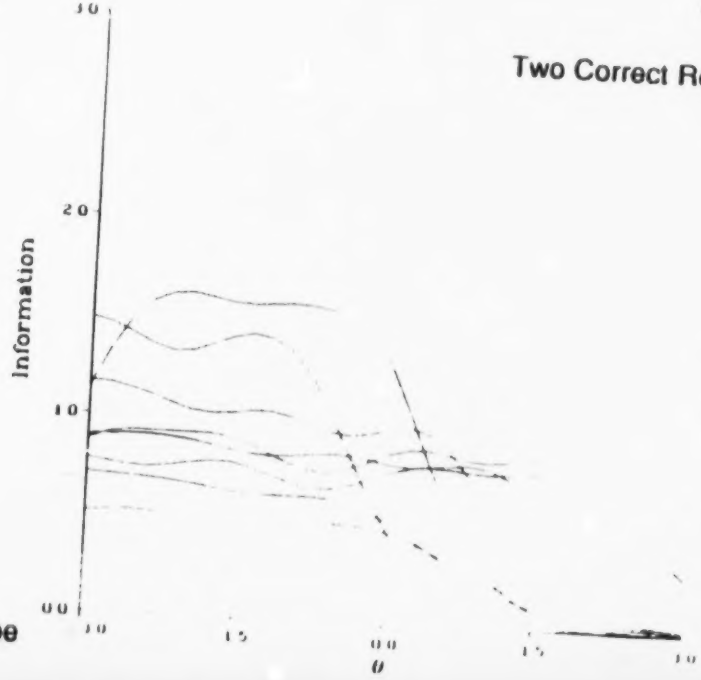
Figure 8



One Correct Response Type

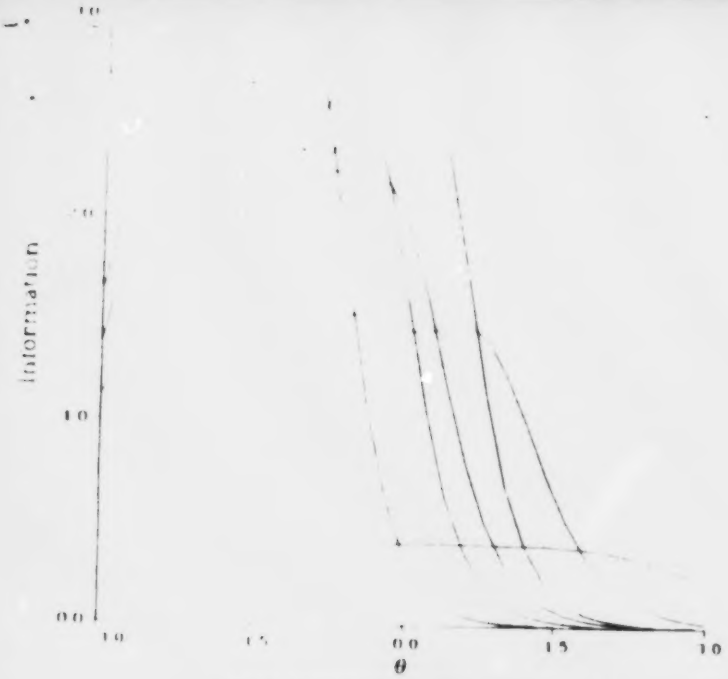


Two Correct Response Type

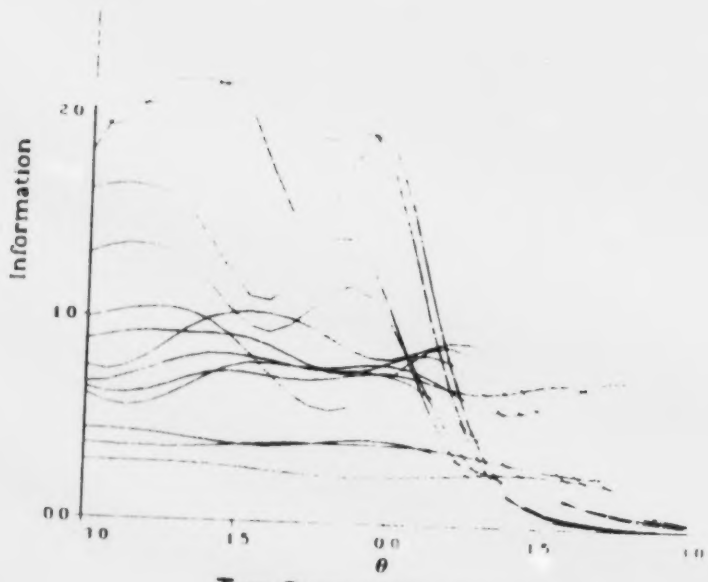


Three Correct Response Type

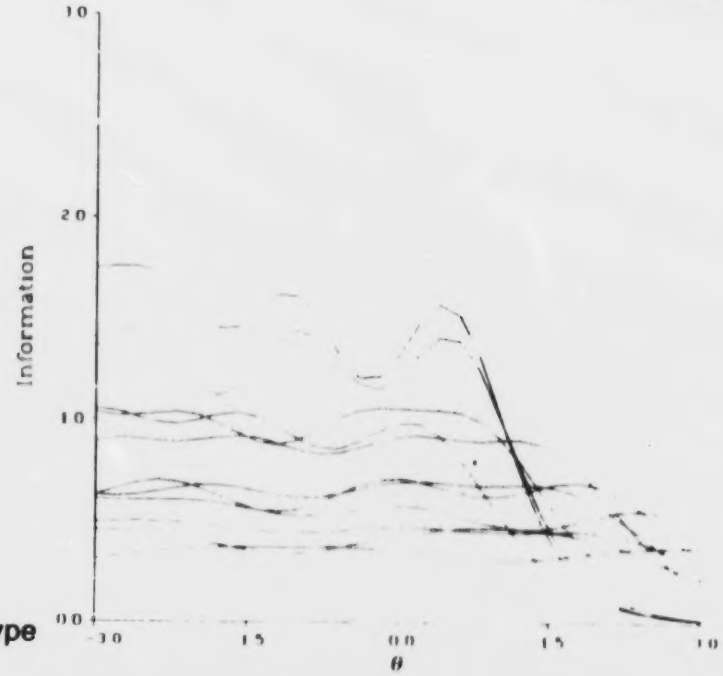
Figure 9



One Correct Response Type

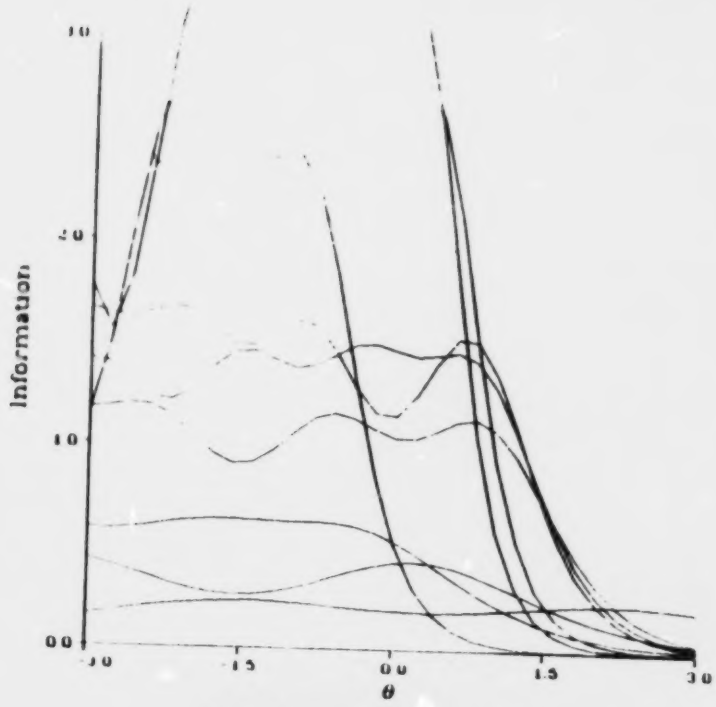


Two Correct Response Type

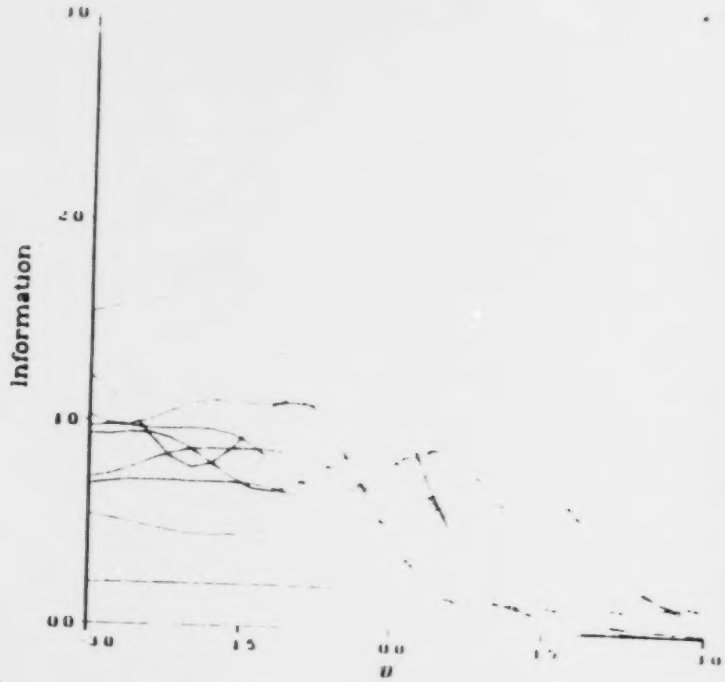


Three Correct Response Type

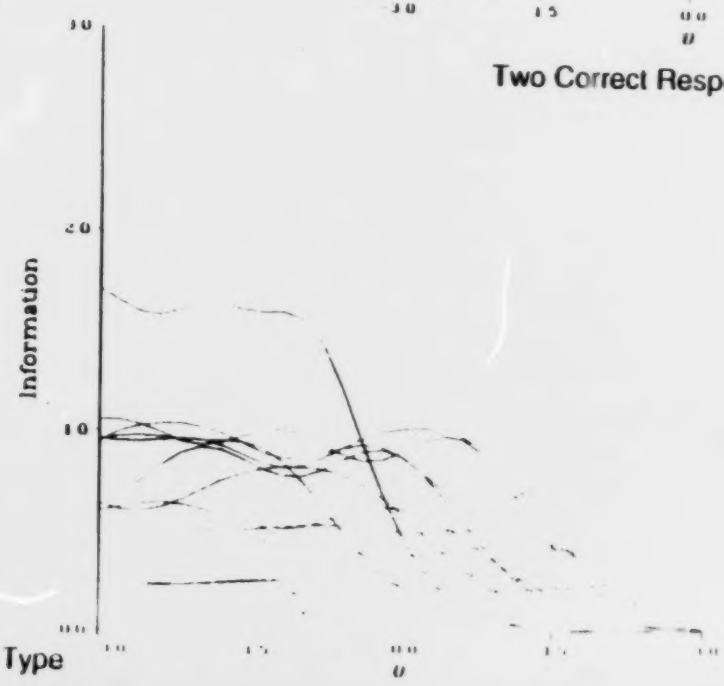
Figure 10



One Correct Response Type

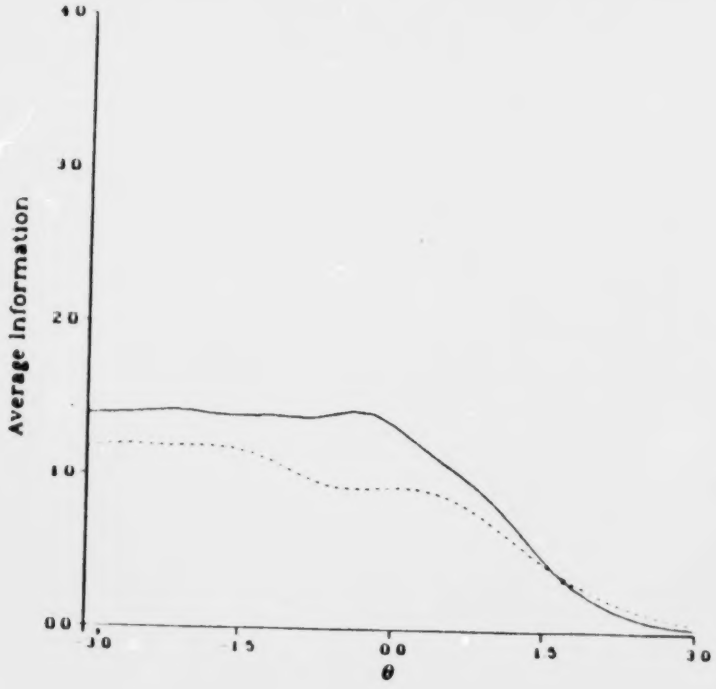


Two Correct Response Type

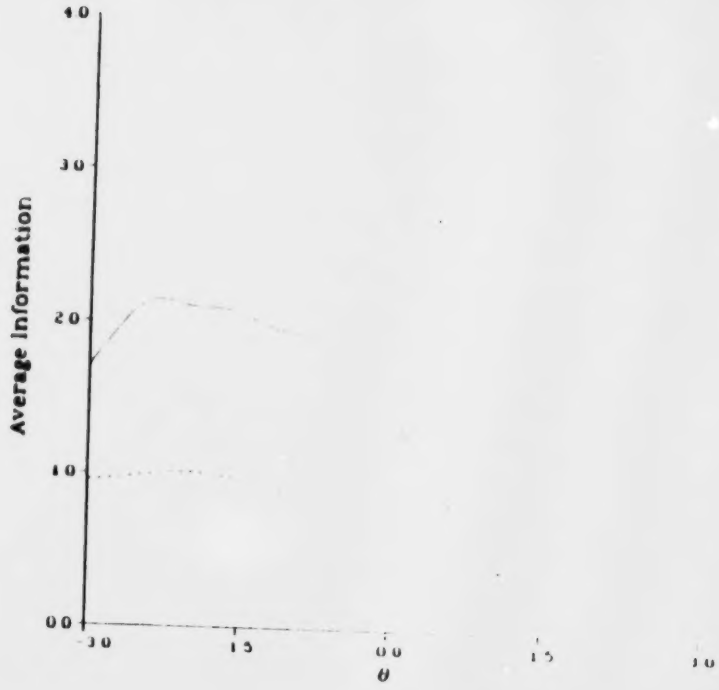


Three Correct Response Type

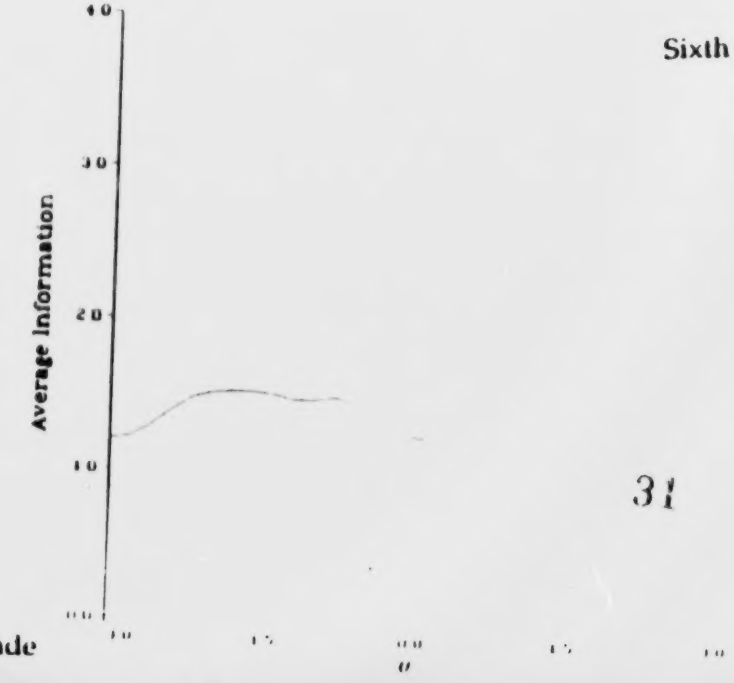
Figure 11



Third Grade



Sixth Grade



Eighth Grade

Sixth Grade IGAP Reading Test